Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP2003/011242

International filing date: 10 October 2003 (10.10.2003)

Document type: Certified copy of priority document

Document details: Country/Office: EP
Number: 03004732.8
Filing date: 04 March 2003 (04.03.2003)

Date of receipt at the International Bureau: 22 December 2006 (22.12.2006)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

EP/03/11242

Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. **Patent application No.** **Demande de brevet n°**

03004732.8

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 03004732.8
Demande no:

Anmeldetag:
Date of filing: 04.03.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

DEUTSCHE THOMSON-BRANDT GMBH
Hermann-Schwer-Strasse 3
78048 Villingen-Schwenningen
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Method and apparatus for describing sound sources

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04N7/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT SE SI SK TR LI

Method and Apparatus for describing sound sources

The invention relates to a method and to an apparatus for describing sound sources, especially for sound sources encoded as audio objects according to the MPEG-4 Audio standard.

Background

The MPEG-4 Audio standard as defined in ISO/IEC 14496-3 and 14496-1 facilitates a wide variety of applications by supporting the representation of audio objects. For the combination of the audio objects additional information - the so-called scene description - determines the placement in space and time and is transmitted together with the coded audio objects.

For playback the audio objects are decoded separately and composed using the scene description in order to prepare a single soundtrack, which is then played to the listener.

For efficiency, the MPEG-4 Systems standard ISO/IEC 14496-1 defines a way to encode the scene description in a binary representation, the so-called Binary Format for Scene Description (BIFS). Correspondingly, audio scenes are described using so-called AudioBIFS.

A scene description is structured hierarchically and can be represented as a graph, wherein leaf-nodes of the graph form the separate objects and the other nodes describes the processing, e.g. positioning, scaling, effects etc.. The appearance and behavior of the separate objects can be controlled using parameters within the scene description nodes.

Empfangszeit 4.März 15:16

Invention

The invention is based on the recognition of the following fact. Currently the MPEG-4 Audio standard cannot describe sound sources that have a certain dimension, like a choir, orchestra, sea or rain but only a point source, e.g. a flying insect, or a single instrument. According to listening tests wideness of sound sources is clearly audible, whereby more complicate descriptions like the shape of the audio object is not necessary.

Therefore, a problem to be solved by the invention is to allow the description of the wideness of sound sources that have a certain dimension in a simple and backwards compatible way.

This problem is solved by the method disclosed in claim 1 and the corresponding apparatus in claim 5.

In principle, the inventive method allows to describe sound sources, which are encoded as separate audio objects. The arrangement of the sound sources in a sound scene is described by a scene description. For playback the audio objects are decoded separately and a single soundtrack is composed from the decoded audio objects using said scene description. For describing the wideness of a sound source an audio spatial diffuseness node is defined within the scene description.

Advantageous additional embodiments of the invention are disclosed in the respective dependent claims.

Drawings

Exemplary embodiments of the invention are described with

reference to the accompanying drawings, which show in

Fig. 1 the illustration of the functionality of the AudioSpatialDiffuseness mode;

Fig. 2 an Audio Scene for a Line Sound Source;

Fig. 3 an exemplary scene with a combination of shapes to represent more complex audio source.

Exemplary embodiments

Figure 1 shows an illustration of the functionality of the inventive AudioSpatialDiffuseness node, in the following also named **AudioDiffusenes node.**

This AudioSpatialDiffuseness node will have a children field as input and will produce the same number of channels (numChan) as output. Branches that are connected to an upper level branch are called children in MPEG-4 terms. It can be inserted in each branch of the audio subtree, without changing any other node.

A diffuseSelection field will allow the scene author to control the diffuseness algorithms, so that each AudioSpatialDiffuseness node will produce a different output. In practice a diffuseness node will virtual produce N different signals, but only one real signal is passed through to the output of the node, signaled by the diffuseSelect field. Other fields like a decorrelation strength (decorrStrength) etc. could be added to the node, if required.

```
AudioSpatialDiffuseness  {
    eventin      MFNode  addChildren
    eventin      MFNode  removeChildren
    exposedField   MFNode  children                    [ ]
    exposedField   SFInt32 diffuseSelect      -1
    exposedField   SFInt32 decorreStrength    1
    field        SFInt32 numChan              1
    field        MFInt32 phaseGroup           [ ]
}
```

Table 1: Semantics of the proposed AudioSpatialDiffuseness Node

In the case of numChan greater than one each channel should be diffused separately.

Figure 2 depicts an Audio Scene for a Line Sound Source. By using this proposal the scene author has to decide how many and at which position the decorrelated multiple point sound sources will be located. The advantage is, that the content author has much more control over the shape effect. He can also use intensity and direction of each point source as well as using the AudioDelay and AudioEffects node for certain Sound nodes to manipulate the effect.

It is still possible for the renderer to reduce the computational power by passing the scene tree to look for identical AudioSources.

```
# Example of a line sound source replaced by three point
sources
# using one single decoder output.

Group {
    children [
```

```
DEF POS1 Sound {
    intensity 0.9
    location 0 0 0
    spatialize TRUE
    source AudioSpatialDiffuseness  {
        numChan 1
        diffuseSelect  1
        children [
            DEF BEACH AudioSource {
                numChan 1
                url 100
            }
        ]
    }

DEF POS2 Sound {
    intensity 0.8
    location -3 0 0
    spatialize TRUE
    source AudioSpatialDiffuseness  {
        numChan 1
        diffuseSelect  2
        children [ USE BEACH]
    }

DEF POS3 Sound {
    intensity 0.8
    location  3 0 0
    spatialize TRUE
    source AudioSpatialDiffuseness  {
        numChan 1
        diffuseSelect  3
        children [ USE BEACH]
    }
]
}
```

Table 2: Example of a Line Sound Source replaced by three Point Sources using one single Audio-Source.

According to a further embodiment primitive shapes are defined and combined using the AudioSpatialDiffuseness nodes to do more complex shapes. An advantageous selection of shapes is e.g. a box, a sphere and a cylinder. All of these nodes should have a location field, a size and a rotation, as shown in table 3.

```
SoundBox / SoundSphere / SoundCylinder {
    eventin      MFNode  addChildren
    eventin      MFNode  removeChildren
    exposedField   MFNode  children              [ ]
    exposedField   MFFloat intensity      1.0
    exposedField   SFVec3f location             0,0,0
    exposedField   SFVec3f size                 2,2,2
    exposedField   SFVec3f rotationaxis         0,0,1
    exposedField   MFFloat rotationangle        0.0
}
```

Table 3

If one size parameter is set to zero a volume will be flat, resulting in a wall or a disk. If two dimensions are zero a line results.

Fig. 3 shows a scene with two audio sources, a choir (or orchestra) located in front of a listener L and audience to the left, right and back of the listener making applause. The choir consists out of one ***SoundSphere*** C and the audience consists out of three ***SoundBoxes*** A1, A2, and A3 connected with ***AudioDiffuseness*** nodes.

A BIFS example for the scene of figure 3 looks as shown in table 4.

```
## The Choir SoundSphere

    SoundSphere {
        location 0.0 0.0 -7.0          # 7 meter to the back
        size 3.0 0.6 1.5             # wide 3; height 0.6; depth
1.5
        intensity 0.9
        spatialize TRUE
        children [ AudioSource {
            numChan 1
            url 1
        }]
    }
## The audience consists out of 3 SoundBoxes

    SoundBox {                       # SoundBox to the left
        location -3.5 0.0 2.0          # 3.5 meter to the left
        size 2.0 0.5 6.0             # wide 2; height 0.5; depth
6.0
        intensity 0.9
        spatialize TRUE
        source AudioDiffusenes{
            diffuseSelect 1
            decorrStrength 1.0
            children [ DEF APPLAUSE AudioSource {
                numChan 1
                url 2
            }]
        }
    }
    SoundBox {                       # SoundBox to the rigth
        location 3.5 0.0 2.0         # 3.5 meter to the right
        size 2.0 0.5 6.0             # wide 2; height 0.5; depth
6.0
```

```
        intensity 0.9
        spatialize TRUE
        source AudioDiffusenes{
            diffuseSelect 2
            decorrStrength 1.0
            children [ USE APPLAUSE ]
        }
    }
    SoundBox {                          # SoundBox in the middle
        location 0.0 0.0 0.0            # 3.5 meter to the right
        size 5.0 0.5 2.0                # wide 2; height 0.5; depth
6.0
        direction 0.0 0.0 0.0 1.0       # default
        intensity 0.9
        spatialize TRUE
        source AudioDiffusenes{
            diffuseSelect 3
            decorrStrength 1.0
            children [ USE APPLAUSE ]
        }
    }
```

Table 4

In this example a children field *APPLAUSE* is defined as an audio source for the first SoundBox and is reused as audio source for the second and third SoundBox. Furthermore, in this case the diffuseSelect field signals for the respective SoundBox which of the signals is passed through to the output.

In the case of a 2D scene it is still assumed that the sound will be 3D. Therefore it is proposed to use a second set of SoundVolume nodes, where the z-axis is replaced by a single float field with the name 'depth' as shown in table 5.

SoundBox2D / SoundSphere2D / SoundCylinder2D {			
eventin	MFNode	addChildren	
eventin	MFNode	removeChildren	
exposedField	MFNode	children	[ ]
exposedField	MFFloat	intensity	1.0
exposedField	SFVec2f	**location**	0,0
exposedField	SFFloat	**locationdepth**	0
exposedField	SFVec2f	**size**	2,2
exposedField	SFFloat	**sizedepth**	0
exposedField	SFVec2f	**rotationaxis**	0,0
exposedField	SFFloat	**rotationaxisdepth**	1
exposedField	MFFloat	**rotationangle**	0.0
}			

Table 5

Claims

1. Method for describing sound sources, which are encoded as separate audio objects, wherein the arrangement of the sound sources in a sound scene is described by a scene description, and wherein for playback the audio objects are decoded separately and a single soundtrack is composed from the decoded audio objects using said scene description, **characterized by** an audio diffuseness node which is defined within the scene description for describing the wideness of a sound source.

2. Method according to claim 1, wherein a diffuse selection field will allow the scene author to control the diffuseness algorithms.

3. Method according to claim 1 or 2, wherein a decorrelation strength field will allow author to control the strenght of the decorrelation.

4. Method according to any of claims 1 to 3, wherein shapes are defined and combined using the AudioSpatialDiffuseness nodes to do more complex shapes.

5. Apparatus for performing a method according to any of claims 1 to 4.

Empfangszeit 4.März 15:16

Abstract

The MPEG-4 Audio standard as defined in ISO/IEC 14496-1 and -3 facilitates a wide variety of applications by supporting the representation of audio objects. For the combination of the audio objects additional information - the so-called scene description - determines the placement in space and time and is transmitted together with the coded audio objects.

For playback the audio objects are decoded separately and composed using the scene description in order to prepare a single soundtrack, which is then played to the listener. A scene description is structured hierarchically and can be represented as a graph, wherein nodes of the graph form the separate objects. The appearance and behaviour of the separate objects can be controlled using parameters within the scene description nodes. For describing the wideness of a sound source an audio diffuseness node is defined within the scene description.

Fig. 1

Empfangszeit 4.März 15:16

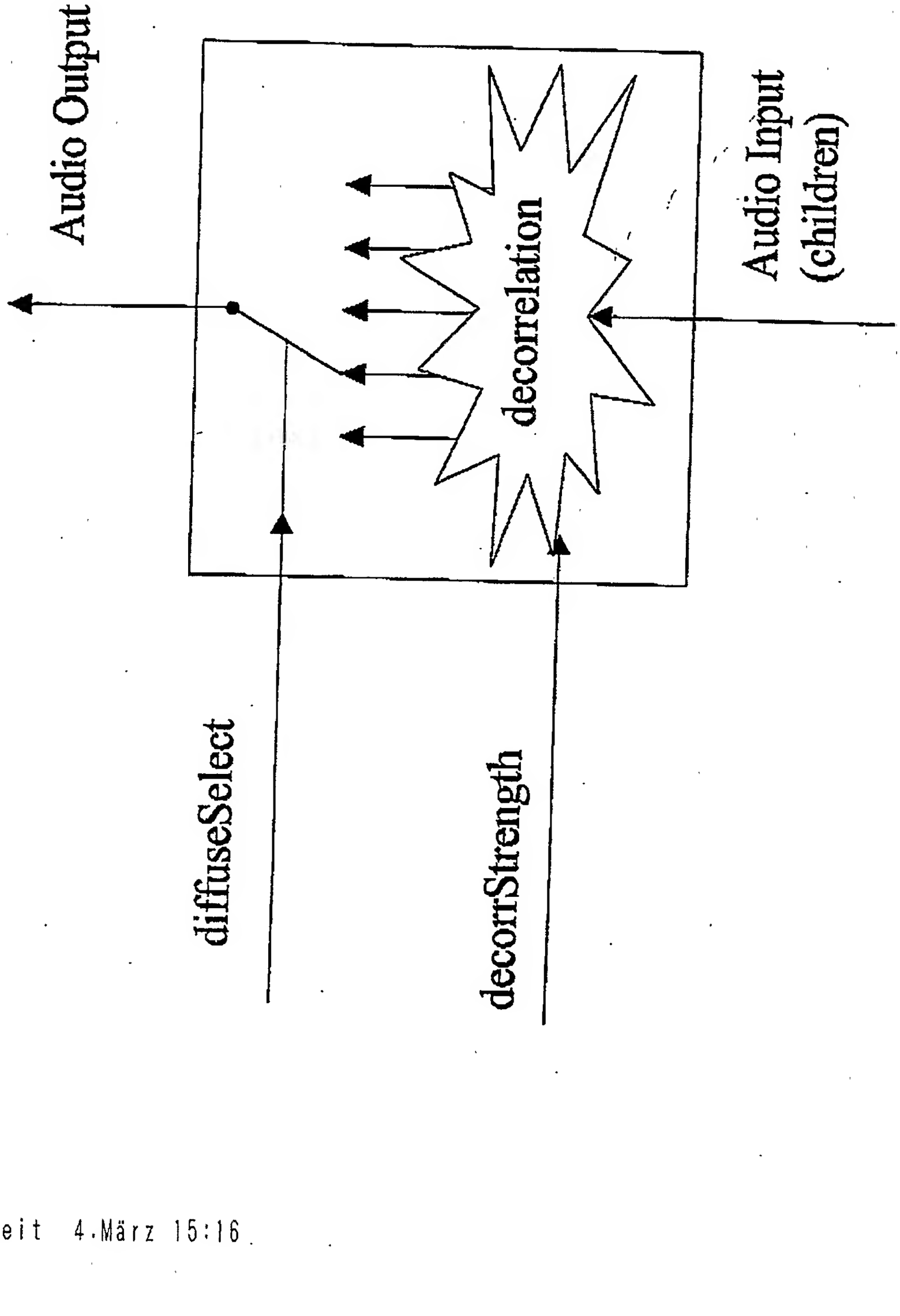
Audio Output
decorrelation
Audio Input
(children)
diffuseSelect
decorrStrength

Figure 1

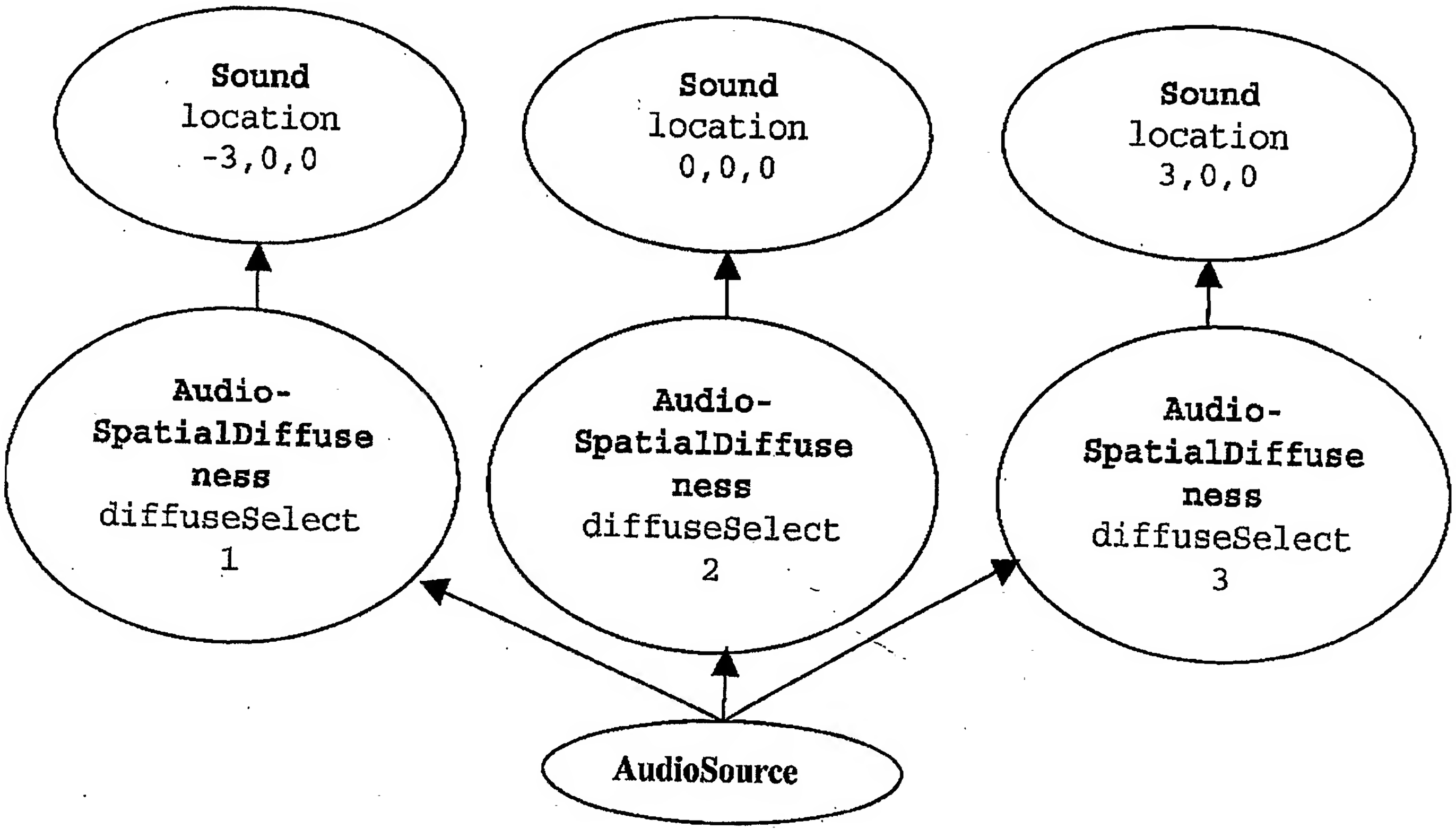

Figure 2

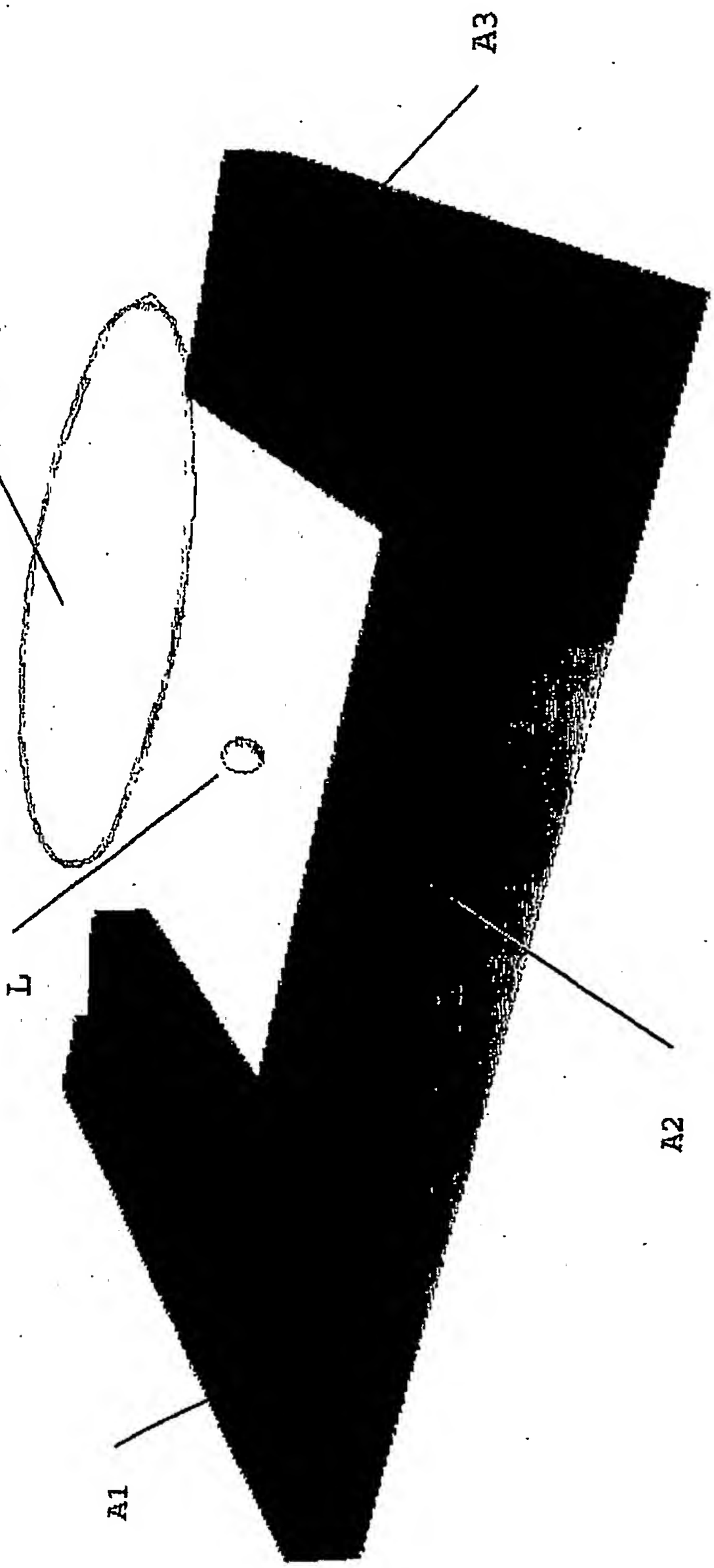

Figure 3